

International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 7.056

IJCSMC, Vol. 14, Issue. 9, September 2025, pg.92 – 102

Intelligent Cloud-Native Data Management on AWS: AI for Migration, Compliance, and Predictive Optimization

Bhargavi Tanneru

btanneru9@gmail.com

DOI: <https://doi.org/10.47760/ijcsmc.2025.v14i09.013>

Abstract: Critical industries, such as telecommunications, financial technology, and healthcare, are undergoing rapid digital transformation, where the shift from legacy systems to cloud-native architectures has become increasingly essential. The complexity of these migrations stems from massive data volumes, stringent compliance requirements, and the need for high availability. Traditional migration approaches often lack the intelligence to automatically classify sensitive data, optimize transformation workflows, or anticipate workload fluctuations. This paper proposes an **AI-driven framework for secure, efficient, and intelligent cloud-native migrations on AWS**, leveraging services such as **Amazon SageMaker, Amazon Bedrock, Amazon Macie, AWS Glue, and Amazon Forecast**. The framework automates data discovery, ensures compliance, and applies predictive optimization to handle large-scale workloads. Through case studies in telecommunications, fintech, and healthcare, we demonstrate how AI reduces downtime, enhances compliance, and achieves cost-optimized scalability. The findings highlight that AI-enhanced migration strategies are not only operationally efficient but also critical in ensuring trust, resilience, and long-term sustainability in data-heavy environments.

Keywords: AWS AI, Cloud-Native Migration, Data Management, SageMaker, Bedrock, Macie, Healthcare IT, FinTech, Telecommunications, Cloud Security

Introduction

The shift toward cloud-native architectures is no longer optional for enterprises operating in **data-intensive and regulated industries**. Telecommunications providers must handle real-time call data and network traffic analytics, while financial technology companies process millions of sensitive financial transactions per second. Healthcare organizations face the challenge of storing, analyzing, and securing electronic health records under **strict compliance mandates, such as HIPAA and PCI-DSS**.

Cloud-native migration involves moving from **legacy monolithic systems** to distributed, containerized, and serverless architectures that run on cloud infrastructure. While AWS provides powerful tools for migration—such as AWS Database Migration Service (DMS), AWS Glue, and AWS DataSync—most of these solutions operate as **reactive, rule-based pipelines**. They require manual oversight, which introduces risks of **downtime, data loss, compliance violations, and unexpected costs**.

Artificial Intelligence (AI), integrated into the AWS ecosystem, provides a transformative opportunity to enhance migration workflows. By embedding machine learning models into **data discovery, schema transformation, and workload optimization**, enterprises can ensure **intelligent, secure, and adaptive migrations**.

This paper explores how AWS AI services can be systematically applied to address key challenges in data-heavy migrations. We propose a **comprehensive AI-driven migration framework**, evaluate its benefits across different industries, and highlight its impact on efficiency, compliance, and cost optimization.

Problem

Despite the growing maturity of cloud migration tools, enterprises continue to face multiple challenges:

1. Downtime and Business Disruption

- Legacy-to-cloud migrations often require batch transfers, leading to **hours or even days of downtime**.
- In mission-critical environments, such as healthcare or telecommunications, downtime directly translates to a **loss of service continuity and customer trust**.

2. Data Sensitivity and Compliance Risks

- Healthcare data (Protected Health Information – PHI) is governed by HIPAA.
- Financial data (credit card numbers, transaction histories) must comply with PCI-DSS.
- Manual classification and compliance validation create **risk of mislabeling and breaches**.

3. Static ETL and Schema Mismatches

- Traditional Extract, Transform, Load (ETL) pipelines are **static** and often fail to adapt to evolving data models.

- Schema mismatches during migration cause bottlenecks and anomalies that **delay cutover timelines**.
4. **Lack of Predictive Intelligence**
- Current migration tools are reactive.
 - Without AI-driven forecasting, organizations face **unexpected workload spikes**, resulting in performance degradation and cost overruns.
- These limitations make **traditional migration strategies insufficient** for industries where **data integrity, compliance, and uptime are non-negotiable**.

Solution: AI-Driven Migration Framework on AWS

To address these challenges, we propose a **multi-layered AI-enhanced migration framework** built on AWS services.

1. Data Discovery & Classification

One of the most critical challenges in migrating enterprise workloads to cloud-native architectures is accurately **discovering and classifying data** before, during, and after migration. In industries such as telecommunications, financial technology, and healthcare, the stakes are exceptionally high because the data handled is often **highly sensitive and subject to stringent regulations**. Examples include **customer call records (telecom), payment card details (fintech), and patient health information (healthcare)**. Failure to correctly identify, classify, and protect this information not only increases the risk of **compliance violations** but also raises the probability of **data breaches** during migration.

Why Data Discovery is Critical

- **Volume & Variety:** Enterprises often have **petabytes of structured and unstructured data** spread across on-premises databases, file systems, and APIs.
- **Hidden Sensitivity:** Sensitive data (e.g., credit card numbers, SSNs, medical notes) may exist in unexpected places such as **log files, spreadsheets, or unstructured documents**.
- **Regulatory Pressures:** Compliance frameworks, such as **HIPAA, GDPR, and PCI-DSS**, mandate stringent controls for identifying and processing sensitive information.

Without automated discovery and classification, migration teams rely on **manual audits** or **schema-based detection**, which are prone to **human error, incomplete coverage, and inefficiency**.

AI-Enhanced Data Discovery on AWS

AWS provides multiple AI-powered services that can **automatically detect, classify, and label data** at scale.

1. Amazon Macie

- A fully managed service that uses machine learning to discover and protect **sensitive data automatically**.
- Identifies **PII, PHI, and financial data** such as Social Security Numbers, credit card details, or medical identifiers.

- Continuously monitors AWS S3 buckets to detect misconfigurations or sensitive data exposure.
- Example: During a healthcare claims migration, Macie can identify files containing **unprotected PHI** and flag them before transfer.

2. Amazon Comprehend

- A Natural Language Processing (NLP) service that applies ML to **unstructured data** (e.g., emails, call transcripts, financial notes).
- Extracts **entities, sentiment, and key phrases** that help in **classifying documents**.
- Example: In telecom, Comprehend can scan customer service call transcripts and categorize them into categories such as **billing, technical issues, or complaints**, ensuring structured storage in the target cloud database.

3. AWS Glue Data Catalog + AI Inference

- Integrates schema discovery with ML-based inference to automatically generate metadata tags.
- Classifies datasets by type, schema, and content, making them easier to govern under **AWS Lake Formation**.

Industry Applications of AI-Driven Data Classification

• Telecommunications:

Call detail records (CDRs) often contain **subscriber identifiers** embedded in log files. AI classification ensures that sensitive identifiers are masked or encrypted during migration.

• Financial Technology:

Credit card data and financial statements must comply with **PCI-DSS**. Macie automatically identifies misclassified or improperly stored data, reducing compliance audit risks.

• Healthcare:

AI models classify **patient records, lab results, and imaging data** to separate PHI from general data. This enables **HIPAA-compliant migration pipelines** where only PHI-tagged data undergoes enhanced encryption and auditing.

Benefits of AI-Based Data Discovery

- **Accuracy:** Machine learning models outperform manual audits in detecting hidden sensitive data.
- **Speed:** Automated scanning reduces **months of manual effort** into **hours or days**.
- **Compliance-Readiness:** Audit-ready reports provide regulators with evidence of proper classification.
- **Continuous Monitoring:** Unlike one-time audits, AI tools provide **ongoing monitoring** even after migration.

Limitations and Considerations

- **False Positives/Negatives:** AI models may misclassify data; hence, **human-in-the-loop validation** may be needed.
- **Cost:** Continuous scanning of petabyte-scale data can become expensive if not optimized.

- **Custom Entities:** Industry-specific sensitive data (e.g., telecom IMSI codes, healthcare ICD-10 codes) may require **custom-trained ML models in SageMaker** to extend beyond Macie’s default detection.

2. Schema Transformation & Anomaly Detection

Once data has been discovered and classified, the next step in migration is **schema transformation**—mapping legacy data models into cloud-native formats such as Redshift, Aurora, DynamoDB, or FHIR-based schemas. Traditional ETL and schema conversion methods are labor-intensive and prone to error, especially when migrating **heterogeneous datasets** from multiple divisions and industries. Schema drift, mismatched encodings, and semantic misalignments often result in costly delays and post-migration failures.

AWS addresses these challenges with AI-driven schema transformation and anomaly detection:

- **AWS Glue** infers schema and automates ETL pipelines.
- **Amazon SageMaker** trains ML models to detect anomalies (e.g., missing attributes, inconsistent codes).
- **Amazon Bedrock** applies LLMs to suggest schema mappings for complex transformations (e.g., COBOL to JSON).

Industry application: Toyota Motor North America (TMNA) – Amazon SageMaker for Data Unification

Toyota Motor North America (TMNA), under the leadership of **Kamal Distell, VP of Data, Analytics, Platforms, and Data Science**, faced the challenge of managing **siload datasets** across connected car platforms, sales, manufacturing, and supply chain operations. These silos hindered the ability to pre-empt quality issues, deliver unified analytics, and build advanced AI applications.

By implementing **Amazon SageMaker**, TMNA established a centralized platform for **data unification and governance**. Key outcomes included:

- **Unified Schemas:** Telemetry, ERP, and supply chain data integrated into a governed cloud-native framework.
- **Anomaly Detection:** ML models identify defects and irregularities in manufacturing data before they escalate.
- **Search & Discovery:** Enterprise-wide access to standardized datasets accelerated cross-team collaboration.
- **Generative AI Readiness:** Unified, governed data provided a foundation for next-generation AI applications.

This case illustrates how AI-driven schema transformation is not just a technical process but a **strategic enabler** of innovation. By leveraging SageMaker for anomaly detection and schema unification, TMNA built a trusted, future-ready data foundation.

3. Predictive Workload Optimization

One of the most critical aspects of cloud-native migration is ensuring that workloads remain **highly available, performant, and cost-optimized**. Traditional migration approaches rely on **reactive scaling**, where systems respond to demand only after spikes occur. While this method

may be sufficient in less dynamic environments, it poses risks in **data-intensive industries** such as telecommunications, financial technology, and healthcare. Sudden surges in data traffic—whether a spike in telecom call records, unexpected market activity in fintech, or seasonal increases in healthcare claims—can overwhelm reactive systems, leading to **downtime, latency, and customer dissatisfaction**.

AI-powered predictive optimization addresses this challenge by enabling systems to **anticipate workload demand and provision resources in advance**. On AWS, this is achieved through services such as **Amazon Forecast, Amazon SageMaker, and Amazon Bedrock**, which integrate machine learning into scaling strategies. Forecast leverages historical workload patterns to anticipate future demand, while SageMaker enables the development of custom predictive models. Bedrock further enhances this ecosystem by using large language models (LLMs) to recommend scaling and optimization strategies for complex workloads. Together, these services enable organizations to shift from a reactive posture to a **proactive, intelligence-driven model**.

Industry application: Healthcare (Anaplan + Amazon Forecast)

Anaplan Inc., a cloud-native enterprise SaaS provider, partnered with AWS to embed **Amazon Forecast** into its **PlanIQ solution**. Healthcare organizations, such as **South Central Ambulance Services (SCAS)** in the UK, deployed this solution in under three weeks, using it to generate **weekly and six-week rolling forecasts** of patient demand. By predicting surges in ambulance requests, SCAS aligned staffing and resource allocation with expected spikes, significantly reducing risk during peak demand. The integration of Forecast enabled SCAS to move beyond reactive staffing and achieve **greater resilience and patient response effectiveness**.

Industry application: FinTech/Retail (The Very Group + AWS Forecast)

The Very Group, the UK's largest integrated digital retailer and financial provider, collaborated with AWS to implement **Amazon Forecast and ML-driven demand planning**. Handling **4.4 million customers** and **1.8 million daily website visits**, the company faced the challenge of managing inventory and financial risk simultaneously. By investing 800+ hours in experimentation and generating over **8 million forecasts**, The Very Group achieved a **9.9% improvement in SKU management**, valued at **£110 million**. Beyond retail, the organization extended Forecast's predictive capabilities into **financial forecasting and credit risk analysis**, showcasing how predictive AI can enhance both **operational supply chain resilience and financial decision-making**.

These case studies demonstrate the versatility of **predictive workload optimization** across sectors. In healthcare, predictive forecasting ensures that resources are ready for emergency services. In fintech and retail, it enables better capital allocation and customer service. By embedding AI into workload management, enterprises can ensure **performance continuity, regulatory compliance, and cost efficiency** at scale.

4. Secure & Compliant Migration

While predictive optimization ensures **system stability and performance**, enterprises in **regulated industries** must also meet stringent **security and compliance requirements** during the migration process. Healthcare organizations must protect **Protected Health Information**

(PHI) under HIPAA; financial institutions must secure **payment card information** under PCI-DSS; and telecom and automotive firms must adhere to **regional data sovereignty and safety regulations**.

Traditional migration approaches often leave security gaps, relying on **manual classification, encryption, and access control**. This reactive model risks **misconfigurations, data leaks, and regulatory violations**. An AI-driven, AWS-native security framework ensures that **sensitive data is continuously protected and governed** during migration.

AWS offers several services that collectively form a **secure and compliant migration stack**:

- **AWS Key Management Service (KMS)**: Provides encryption at rest and in transit for sensitive datasets.
- **AWS Identity and Access Management (IAM) and Lake Formation**: Enforce least-privilege access policies across teams and datasets.
- **Amazon Macie**: Uses machine learning to detect and classify sensitive data, such as PII or PHI.
- **AWS CloudTrail**: Maintains audit-ready migration logs for regulators.
- **Amazon SageMaker**: Supports building compliance-specific anomaly detection models for continuous governance.

Industry Application: Robinhood – FinCrimes Agent with Amazon Bedrock

Robinhood, a leading financial services provider, developed its **FinCrimes Agent**, a multilayered AI-driven system designed to streamline investigations into potential financial crimes. Built on **Amazon Bedrock** and orchestrated with **Amazon RDS**, the solution uses multiple **large language models (LLMs)** for tasks such as summarization, classification, validation, and synthesizing external data.

When suspicious activity is detected, the FinCrimes Agent orchestrates specialized sub-agents through an asynchronous task queue. These agents analyze both **structured and unstructured data** from Robinhood's databases and public sources, generating concise investigative summaries that save analysts significant time. The solution employs multiple LLMs, including Anthropic's Claude and DeepSeek, selecting models dynamically to optimize cost and accuracy. Key security and compliance measures were built into the workflow:

- **Virtual Private Cloud Deployment**: Models run inside Robinhood's VPC on Amazon Bedrock, ensuring sensitive data never leaves company control.
- **Explainability and Validation**: Every AI-generated output is paired with a validation agent that checks for factual accuracy and coherence, applying hallucination detection before results are passed forward.
- **Auditability**: Benchmarking against resolved cases ensures continuous monitoring, while audit logs provide regulators with verifiable evidence of compliance.
- **Human-in-the-Loop**: Investigators retain accountability for final decisions, balancing automation with regulatory responsibility.

The impact has been significant:

- **20% efficiency gain** in investigative workflows.

- Faster annotation and reporting, reducing time spent collecting and summarizing data.
- Improved accuracy in identifying illicit activity through AI-assisted workflows.

By aligning AI-driven automation with **regulatory standards for security, explainability, and governance**, Robinhood has set a **new benchmark in financial crime investigation**. The company's FinCrimes Agent demonstrates that **responsible generative AI**, when embedded in a secure AWS environment, can accelerate compliance-driven operations without compromising trust.

Impact

The adoption of an **AI-enhanced cloud-native migration framework** on AWS produces impact that goes beyond technical acceleration. It reshapes how enterprises **govern data, optimize performance, maintain compliance, and unlock business value** across industries. The following dimensions of impact are observed:

1. Operational Efficiency and Downtime Reduction

- AI-driven automation reduces manual overhead in discovery, schema transformation, and workload scaling, compressing migration timelines from weeks to days.
- Example: **South Central Ambulance Services**, through Anaplan's PlanIQ and Amazon Forecast, deployed forecasting models in under three weeks, enabling resource planning for patient surges without disrupting emergency services.

2. Data Quality, Accuracy, and Anomaly Detection

- Machine learning ensures schema consistency and prevents propagation of errors during migration.
- Example: **Toyota Motor North America (TMNA)** unified siloed automotive datasets with SageMaker, identifying anomalies in manufacturing data early and improving cross-enterprise governance.

3. Compliance Assurance and Regulatory Trust

- AI-based classification, encryption, and auditability strengthen compliance with HIPAA, PCI-DSS, and GDPR.
- Example: **Robinhood's FinCrimes Agent**, powered by Amazon Bedrock, integrates validation agents, hallucination checks, and audit logs to meet explainability requirements while improving investigative workflows.

4. Financial Optimization and Business Value Creation

- Predictive scaling and optimized resource allocation reduce infrastructure costs while improving ROI.
- Example: **The Very Group** achieved a **9.9% improvement in SKU management**, generating more than **£110 million in value**, by applying Forecast-based demand prediction at retail scale.

5. Innovation Enablement and Future-Readiness

- Unified, AI-governed data serves as a foundation for generative AI applications, predictive analytics, and autonomous cloud operations.
- Example: TMNA's unified data platform not only improved quality control but also positioned the enterprise to develop **generative AI solutions** for predictive maintenance and customer experience.
- In parallel, **NFL Next Gen Stats**, powered by AWS, demonstrates how predictive and interactive AI can transform **fan engagement and media coverage**, extending impact beyond regulated industries into large-scale consumer experiences.

Collectively, these outcomes demonstrate that AI-driven cloud-native migration is not a narrow technical enabler but a **strategic driver of digital transformation**. By embedding intelligence into every stage of the migration lifecycle, organizations can achieve **resilience, compliance, and scalability** while unlocking new opportunities for **innovation and growth**.

Scope

The scope of this paper extends across both **technical and industry dimensions**, reflecting the versatility of AI-driven cloud-native migration frameworks on AWS. While the case studies presented highlight healthcare, financial technology, automotive, and telecommunications, the principles and framework described are broadly applicable to any industry facing the dual challenge of **data-heavy workloads and strict regulatory requirements**.

1. Industry Coverage

- **Healthcare:** AI-driven classification and compliance mechanisms ensure HIPAA-aligned migration of electronic health records and claims data.
- **Financial Technology & Retail:** Predictive forecasting and compliance workflows optimize both operational efficiency and regulatory trust, as demonstrated by Robinhood and The Very Group.
- **Automotive:** Schema unification and anomaly detection create a governed data foundation for predictive quality management and future generative AI applications.
- **Telecommunications:** Predictive scaling and anomaly detection support real-time call data management and network optimization.
- **Sports & Entertainment (Emerging):** NFL's Next Gen Stats demonstrates how AWS AI services extend to fan engagement and large-scale event analytics, underscoring the broad applicability of the framework.

2. Technical Scope

- The framework integrates AWS services for **data discovery (Macie, Comprehend), schema transformation (Glue, SageMaker, Bedrock), predictive optimization (Forecast, Bedrock), and compliance (KMS, IAM, CloudTrail)**.
- While focused on AWS-native implementations, the architectural principles are adaptable to **hybrid and multi-cloud environments**.

3. Future Directions

- **Autonomous Compliance:** Integration of AI into DevSecOps pipelines for continuous compliance monitoring across regulated industries.
- **Generative AI Integration:** Use of foundation models on Amazon Bedrock to enable advanced applications such as intelligent assistants for compliance, predictive maintenance, and customer engagement.
- **Multi-Cloud Portability:** Expansion of the framework beyond AWS to ensure interoperability across **Azure, GCP, and hybrid environments**, supporting organizations with diverse cloud strategies.
- **Industry-Specific Extensions:** Tailoring AI migration models to sector-specific standards (e.g., FHIR for healthcare, PCI-DSS for finance, ISO 26262 for automotive).

Overall, the scope of this work is not limited to migration itself but encompasses the **full lifecycle of cloud-native adoption**, where AI enhances not only the **movement of data and applications** but also their **governance, optimization, and future innovation potential**.

Conclusion

The migration of enterprise workloads to cloud-native architectures has become inevitable, particularly in industries where **data intensity, compliance, and performance resilience** are non-negotiable. Traditional approaches—manual ETL pipelines, reactive scaling, and rule-based compliance checks—fall short in addressing the complexity of today’s enterprise environments.

This paper presents an **AI-driven framework for cloud-native migrations on AWS**, demonstrating how services such as **Amazon SageMaker, Amazon Bedrock, Amazon Forecast, Amazon Macie, and AWS Glue** can be orchestrated to deliver **intelligent, adaptive, and secure migrations**.

Through real-world case studies, we have shown how:

- **Healthcare providers** (e.g., South Central Ambulance Services) improve patient response by forecasting demand surges with Amazon Forecast.
- **Financial and retail enterprises** (e.g., The Very Group, Robinhood) achieve both compliance and efficiency, from SKU management worth over £110M to AI-augmented financial crimes investigations.
- **Automotive leaders** (Toyota Motor North America) unify siloed data with SageMaker, laying the foundation for generative AI.
- **Sports organizations** (NFL Next Gen Stats) extend AWS AI into fan engagement, proving scalability beyond regulated industries.

The collective evidence suggests that AI-enhanced migrations are not merely a **technical upgrade**, but a **strategic enabler of digital transformation**. Organizations achieve reduced downtime, improved compliance assurance, cost efficiency, and—most importantly—**future readiness for generative AI and autonomous cloud operations**.

Looking forward, the trajectory of AI-driven cloud migrations points toward **self-healing, continuously optimized infrastructures**, where compliance, workload management, and innovation are embedded into the fabric of cloud-native systems. AWS, with its integrated AI services and global adoption, is uniquely positioned to power this transformation across industries.

In conclusion, AI-powered migration frameworks are redefining not only how enterprises move to the cloud but also how they **compete, comply, and innovate** in the digital era.

References

- [1]. Amazon Web Services, “Amazon SageMaker – Build, train, and deploy ML models at scale,” AWS, 2025. [Online]. Available: <https://aws.amazon.com/sagemaker/>
- [2]. Amazon Web Services, “Amazon Macie – Discover and protect sensitive data,” AWS, 2025. [Online]. Available: <https://aws.amazon.com/macie/>
- [3]. Amazon Web Services, “Amazon Bedrock – Build and scale generative AI applications,” AWS, 2025. [Online]. Available: <https://aws.amazon.com/bedrock/>
- [4]. Amazon Web Services, “AWS Glue – Serverless data integration,” AWS, 2025. [Online]. Available: <https://aws.amazon.com/glue/>
- [5]. Amazon Web Services, “Amazon Forecast – Accurate time series forecasting,” AWS, 2025. [Online]. Available: <https://aws.amazon.com/forecast/>
- [6]. M. Armbrust et al., “A view of cloud computing,” *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.
- [7]. R. Buyya, J. Broberg, and A. Goscinski, *Cloud Computing: Principles and Paradigms*. Wiley, 2011.

- [8]. Gartner, "Cloud Migration Strategies: Best Practices," Gartner Research, 2024.
- [9]. G. S. Panesar and R. Chadha, "Optimizing Cloud Environments: Machine Learning-Driven Virtual Machine Migration Strategies," *2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, Tashkent, Uzbekistan, 2023, pp. 1455-1462, doi: 10.1109/ICTACS59847.2023.10389990.